SURVEY SAMPLING

SURVEY SAMPLING

Leslie Kish
Professor of Sociology
Program Director, Survey Research Center
Institute for Social Research
The University of Michigan

JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto

311.21
K61s
cop. 2

AL

Copyright © 1965 by
John Wiley & Sons, Inc.

All Rights Reserved.
Reproduction or translation of any part of this work beyond that permitted by Sections 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

12 13 14 15

Library of Congress Catalog Card Number: 65-19479
Printed in the United States of America

ISBN 0 471 48900 X

Preface

This is designed to be a simple book on sampling methods, with emphasis on and illustrations from surveys of human populations. Although I have designed samples for a variety of nonhuman populations, my experience has been mostly in social applications; the book emphasizes these areas, and it draws most of its examples from them. However, essentially the same methods can be used for sampling animals, plants, minerals, physical products, accounts, and inventories. Sampling specialists find that their skills possess broad generality and transferability; thus this book can also be useful to medical men, biologists, chemists, engineers, and accountants.

Sampling plays a vital role in research design involving human populations; it commands increasing attention from social scientists and practitioners. I include not only the disciplines of economics, sociology, anthropology, psychology, and political science, but also the professions of public health, biostatistics, education, social work, public administration, and business administration. Sampling problems are equally material to practitioners engaged in marketing, commerce, and industry.

While writing this book, I tried to keep a definite audience in mind and to maintain a dialogue with them. My first consideration was for quantitatively oriented students in the social sciences and allied fields. I believe that sampling methods should become roughly the third course of statistics for students in the social sciences. It should follow a course in the fundamentals of statistical reasoning, and another devoted to major statistical tools. I hope that this book can serve as a text for professors of social statistics who are not specialists in sampling.

Second, I have also aimed to provide a reference book on sampling

UNIVERSITY LIBRARIES
CARNEGIE-MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213

SEP 21 '81 SEP 21 '81

methods for professors, researchers, and officials who want to understand survey sampling without necessarily becoming sampling specialists. Improving their methods is crucial, since most samples in survey research represent occasional small or medium-sized efforts by researchers who are not specialists.

I have not written this for my colleagues in sampling, but chiefly for my fellow social scientists, whose specialties differ from mine. Reading one book will not turn anyone into an expert sampling statistician; but the careful reader should become able to design and execute valid samples of moderate dimensions and difficulty, to avoid selection biases, and to achieve reasonable efficiency. He should also become more adept at evaluating the sample results he encounters, to judge their validity, their limits of inference, applicability, and precision.

The reader will be able to design satisfactory rather than optimum samples—to *satisfice* (as Herbert Simon says), rather than to optimize. I paid less attention than others to some technical efforts for getting the last 2 percent efficiency out of samples. These efforts may be worthwhile for large, complicated samples—where it may be best to consult an experienced sampling statistician. In that case, the researcher with sufficient knowledge to meet him half-way will be better able to utilize the scarce time of the sampling consultant.

Social research places special emphasis on the comparative and analytical uses of samples. Hence, I have stressed statistics of subclasses in the sample, rather than focusing all attention on the entire sample. Subclasses are treated for each of the major designs; and the comparison of subclass means receives frequent attention. Other issues relating to analytical statistics from complex samples are also discussed. These discussions frequently utilize research developed under a grant (G-7571) from the National Science Foundation.

The book is oriented toward providing a working knowledge of practical sampling methods, with an understanding of their theoretical background. The necessary working formulas are given. I have tried to state clearly the assumptions underlying the formulas, and to outline the potentialities and limitations inherent in the assumptions. I often use illustrations to explain in detail the meaning of formulas and definitions. I also provide a variety of examples, with computations laid out in painstaking detail.

The book contains many practical procedures, the "domestic arts" of sampling along with its science; the valuable "tricks" that are usually learned only in apprenticeship. Perhaps the principal task of any book or course on methods should be the drastic shortening of the period of apprenticeship in a discipline. Throughout the book the reader can

discover my attempts to synthesize complete sets of coherent rules for conduct ordinarily explained—if at all—with *ad hoc* rules-of-thumb.

My efforts are perhaps most evident in methods for dealing with frame problems. They receive a great deal of emphasis, especially in an early summary (2.7) of frame defects and techniques for treating them, and in a detailed treatment in Sections 11.1 to 11.4. Throughout the book the reader's attention is called to possible frame defects and their effects on sample design. Problems invite the reader to participate in finding and overcoming defects he is likely to encounter in research situations.

Sampling methods are not developed here for their own sake, but as means to ends originating in substantive research problems, especially in the social sciences and their applications. They are designed to provide valid, scientific, and economical tools for those research problems. Hence, the approach is directed less toward delivering a neat theoretical package than toward practical research projects in all their complexity, while preserving a fundamental simplicity of presentation.

To achieve simplicity I frequently resort to useful approximations. Briefly, this book attempts to preach what reasonable sampling statisticians actually practice, rather than an arid dogma. For example, the factor $(N - 1)/N$, if inconvenient, can be ignored when the population size N is large. The book emphasizes estimated variances that the researcher actually computes, rather than their expected values. And so on.

I did not try to make this book "self-contained"—an illusory aim. For the algebraic steps of derivations the interested reader is directed, by frequent references, to relevant passages in several fine textbooks. In these references I sought neither to discover original sources, nor the prestige of the most elegant mathematical treatment. Instead, I sought the most relevant, readable, and available sources, preferably textbooks; there the reader can find further references.

However, I am glad that, on the urging of friends, I abandoned my original intention to avoid derivations completely. My derivations are few and simple, yet they cover the essentials of sampling theory adequately for readers who understand certain fundamental concepts of statistical theory. Sampling theory was condensed by eliminating some repetition and detail. The basic derivations are in Sections 2.8, 4.5, 4.6, 5.6, 6.6, 8.5, and in scattered remarks. These and other technical portions carry a (*); continuity can be maintained without them.

My original intention was to formalize and disseminate the notes that my students and I have evolved from teaching a one-semester course. This restricted but vital core is presented in the simple, nontechnical, carefully organized and related sections of Chapters 1 to 8 in Part I: Fundamentals of Survey Sampling. To complete the course, I add

Sections 9.1 to 9.5 on area sampling, plus most of Chapter 13 on non-sampling errors. Brief selections may also be added from Chapters 11 and 12, which deal with special techniques.

These two chapters and the technical portions of others have served as the base for advanced courses in special problems. However, they are designed primarily to serve as sources of reference for individual, technical problems. These sections are mostly self-contained, rather than interconnected. I hope that researchers will often find in one of them ready answers to specific technical problems.

Those who can avoid the complexities of cluster sampling will find a self-contained treatment of element sampling in Chapters 1 to 4. To this can be added a modest treatment of the essentials of cluster sampling: 6.4B, 6.5A and B, 7.2, 7.3, and 9.1 to 9.3. These could comprise a third of a course, incorporated into a full year course in statistics.

For a briefer introduction to sampling, teachers have used the following sections of 81 pages: 1.0–1.7, 2.6, 2.7, 3.1, 3.4, 3.5, 5.1, 5.4, 6.4B, 6 5A, 7.2, 7.3, 9.1–9.3.

I hope that readers will especially like the problems. They should give much credit to my students, who eliminated the worst, stimulated the best, and sharpened their phrasing; also some credit to my many colleagues and clients, whose questions and research projects gave birth to most of these problems. I tried hard to make them life-like; hence, the student's crucial task is to discover what the principal issues are, then pose appropriate questions. When this is well done, the answers come rather smoothly. The focus is on principal issues, rather than on fine nuances. This was made possible by eliminating cumbersome details, while retaining the essence of actual problems. In addition to testing the student's ability to apply the book's methods, these problems perform a double task. They deepen understanding of those methods by extending them to new and varied areas. Furthermore, the extensions themselves represent valuable methodological tools. Work on these problems should yield half the value the student can extract from this book. Of a subject devoted to methods, we can say with Sophocles: "One must learn by doing the thing; for though you think you know it, you have no certainty, until you try."

It is with genuine pleasure that I acknowledge my debt to many friends and colleagues. Since Yates' book in 1948, six fine books on sampling have appeared in English, and I have learned and taught from each in turn. Deming, Cochran, Hansen, and Hurwitz have been personal teachers, friends, and consultants. I was fortunate to have, at different times, Roe Goodman, Benjamin Tepping, and Irene Hess as colleagues in the Sampling Section of the Survey Research Center. Angus Campbell and Rensis Likert, in creating and directing that Center and the Institute

for Social Research, have provided valuable support and opportunity. Questions, problems, and issues posed daily by the social scientists at the Center provide a uniquely stimulating atmosphere. I also benefited from daily contact with colleagues in the Sociology and Mathematics Departments. Of the many statisticians who have given me personal stimulation, support, and criticism, I must single out L. J. Savage, Bruce Hill, and Howard Raiffa; William Ericson gave me valuable criticism on much of the final draft, tightening many a loose phrase or formulation. They stimulated and helped me to make this book more suitable as a text in departments of statistics than I first planned. V. K. Sethi and C. T. Tharakan eliminated many technical errors. Above all, I was fortunate to find in my wife, Rhea, an expert and devoted editor, who has greatly improved the book's style.

Primary and essential stimulation came from the many students in about forty courses in sampling I have given in fifteen years. They represent a good cross section of researchers, teachers, practitioners, and students from the social sciences, mathematics, and allied fields. The students in the seventeen yearly Summer Institutes on Survey Techniques have been mostly officials, scientists, and professors; they came from all over the globe, representing yearly an average of ten countries. Their experience, searching questions, suggestions, and arguments have contributed substantially to the origin, form, and content of this book.

Leslie Kish

Ann Arbor, Mich.
June 1965